

A MetaMorpho fordítóprogram projekt 2005-ben

Tihanyi László

MorphoLogic
1126 Budapest, Orbánhegyi út 5.
tihanyi@morphologic.hu

Kivonat: Az előadásban felsorolom a MetaMorpho fordítóprogram projekt 2005-ös évben elért eredményeit a MorphoLogicon belül. Röviden ismertetem a MorphoLogic, a Nyelvtudományi Intézet és a Szegedi Egyetem alkotta konzorcium keretében folyó magyar–angol fordítóprojekt első évében történt eseményeket. Beszámolok a MoBiCAT internetes fordítószolgáltatás elsőéves tapasztalatairól. Ez a cikk folytatása a 2003-as és 2004-es MSZNY-konferencián elhangzott projekt beszámolóknak [1],[2], és összefoglalja a 2005-ös év legfontosabb fejlesztéseit.

1 Programfejlesztések

1.1 Január

Megoldottuk a birtokos szerkezetek felhasználói szótárba történő felvételét. Az egyértelműség biztosítása érdekében kötelezővé tettük a genitivus jelölését, és egy külön beállítással biztosítottuk, hogy az eredményben jelöletlen maradjon.

Formátumkonverzió a mintákhoz. A magyar-angol fordítóprogram fejlesztés keretén belül felmerült az igény a MetaMorpho projektben használt belső mintaformátum (MMD) szabványos XML alakra történő konverziójára.

Karakter-alapú elszámolás. A MoBiCAT szolgáltatás havi előfizetéses rendszerben működik. Erre az ad lehetőséget, hogy a felhasználó egyszerre csak egy-egy mondat fordítását igényelheti. A teljes szöveg fordítását megengedő, böngészőből vagy Wordből működtethető fordítói szolgáltatások díjazásának a fordított szöveg mennyiségével kell arányban állnia. A kifejlesztett karakterszámlálás a forrás vagy a célnyelvi karakterek száma alapján is működtethető.

A RuBi mintabővítő eszközünkben megjelent a szakterület- (vagy domain-) kezelés. Szabadon definiálhatók új szakterületek, és különböző területeken egy adott szóhoz különböző jelentéseket rendelhetők. Fordításkor a megfelelő szakterület választása mellett a szó a kívánt jelentésével fog szerepelni.

Beindult a MetaMorpho tesztablaka a www.metamorpho.hu oldalon. Az MmoText egy olyan tesztfelület, amellyel korlátozott hosszúságú angol szövegek fordíthatók le magyarra. A felület ingyenes, és azonnali kipróbálására ad lehetőséget. A felület

kódszamos védelemmel van ellátva, így a fordítandó szöveggel együtt egy csak emberek által értelmezhető számsorozat is meg kell adni.



MetaMorpho

Nyelv:

This is the test window of the machine translation program.

Ez a gépi fordítóprogram tesztablaka.

1.2 Február

Elválasztottuk a belső és a termékbe szánt mintabővítő változatokat. Egy jól kitöltött minta létrehozásához szakmai képzettségre volt szükség, ezért ezeken egyszerűsítettünk. Ugyanakkor belső használatban ezeknek a tulajdonságoknak a helyes beállítását továbbra is biztosítani akartuk.

Bevezettük a kompatibilitási operátort, amely a minták típusának közvetlen levizsgálhatóságán keresztül a minták számának, és így a nyelvtan méretének csökkenéséhez vezetett.

Megszületett az első magyar–angol minta konverter. A fejlesztésben csak konverternek hívott eszköz, a megírt szabályokat alakítja át egy olyan még mindig a megírt szabályokkal azonos formátumú forrássá. A konverter valósítja meg pl. az öröklési mechanizmust, de a metatulajdonságok segítségével a nyelvtan írója tetszőleges átalakításokat és kiegészítéseket végezhet.

Megvalósítottuk az előfizetők saját webes adminisztrációs felületét. A regisztrált felhasználók ezen a felületen ellenőrizhetik pl. különböző fordítószolgáltatásaik állapotát.

A MorphoWord Pro változatát egy USB-portba helyezhető hardverkulcsos védelemmel láttuk el.

Elértük, hogy a felhasználói minták mindig érvényre jussanak, azaz felülbírálják a rendszerbe épített tudást.

Nyelvileg értékeltük a szolgáltatás közben létrejövő fordításokat, és a gyakori ismeretlen szavakkal kiegészítettük a szótárt.

Beindult a hamarosan nyilvánosan is elérhető lesz a MorphoWeb weblapfordító. A programtól tetszőleges angol internetes oldalt fordítása kérhető. A program követni tudja a linkeket is, így a fordított oldalról továbblépve további fordított oldalakra juthatunk.

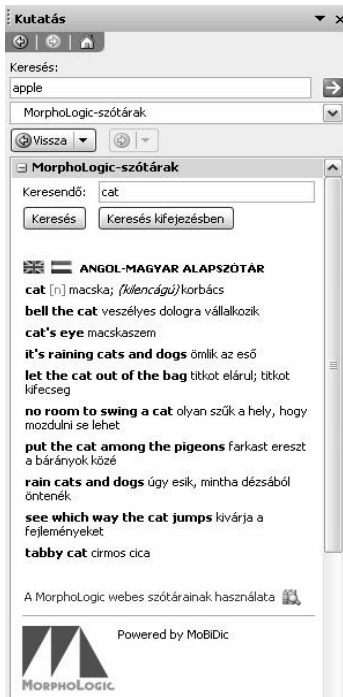
1.3 Március

A programot összehasonlítottuk más fordítóprogramokkal. A vizsgálathoz angol-német fordítókat használtunk, úgy hogy az angol szövegekhez három német és három magyar emberi fordítást készítettünk. A vizsgálat alapján megállapítható volt, hogy a legjobb eredményt az orosz PROMT érte el. A MetaMorpho a piacon létező fordítókkal gyakorlatilag azonos minőséget produkált. Kiemelendő, hogy a leghosszabb szó szerinti egyezést (12 szó) a MetaMorpho érte el.

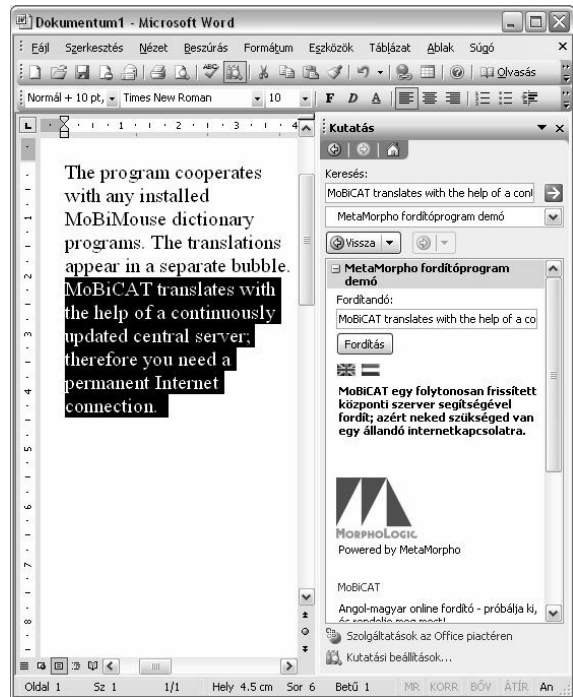
	Systran	SDL	PROMT	Wordlingo	MMO
Bleu (1-4 gram)	0,4206	0,4138	0,4618	0,4158	0,4101
1 gram	0,6519	0,667	0,6848	0,6528	0,6230
max gram	10	7	8	10	12

Megvalósult a Microsoft Office 2003 szövegszerkesztőből a Kutatási felületen keresztül működtethető ingyenes angol-magyar és magyar-angol szótár, valamint az angol-magyar fordítóprogram demó.

Office 2003 Research angol-magyar szótár



Office 2003 Research fordítóprogram-demó



1.4 Április

Megvalósult a vonzatos és vonzat nélküli igék egységes kezelése az adatbázisban.

Az Európai Unió sikeres vizsgálatot folytatott a MetaMorpho programról. A részletes felmérés eredményeként a programot nemcsak az Uniós magyarról és magyarra fordításokhoz javasolják, de alkalmasnak találták további uniós nyelvpárok fejlesztésére is.

1.5 Május

A mySQL mellett már a szabadon termékbe építhető SQLite adatbázis-kezelővel is építhetők a felhasználói szótáraink.

Létrejötték a MorphoWord termék telepítőprogramjai.

1.6 Június

Összehangoltuk a szintaktikai és morfológiai szótárainkban lévő szavakat, és kiegészítettük az adatbázisokat.

Elkészültek a MorphoWord program dokumentációi.

1.7 Július

A bálna (MoBiDic), az egér (MoBiMouse) és a macska (MoBiCAT) után megszületik a negyedik „MorphoLogic-állat”: a MorphoWord lepke, mely nevét a dél-amerikai kék Morpho lepkétől (is) kölcsönzi.

Elkészül a MorphoWord aktiválós szoftvervédelme.

A MoBiCAT az Internet Explorer mellett már Mozilla FireFox alatt is működik.

Egy felmérés készül a MoBiCAT-vásárlókról és szokásaikról. A felmérésben a szokatlanul magas érdeklődés mellett hasznos információkhoz jutunk a további fejlesztésekhez.

1.8 Augusztus

Reklámakciót szervezünk a Sziget-fesztivál alatt.

Kiegészül a MorphoWord telepítőprogramja, amely a MorphoLogic-szótárak demó változatai mellett az xPlace intelligens szócserélőt is tartalmazza.

Kialakul a MorphoWord designja.

Tesztelések folynak.

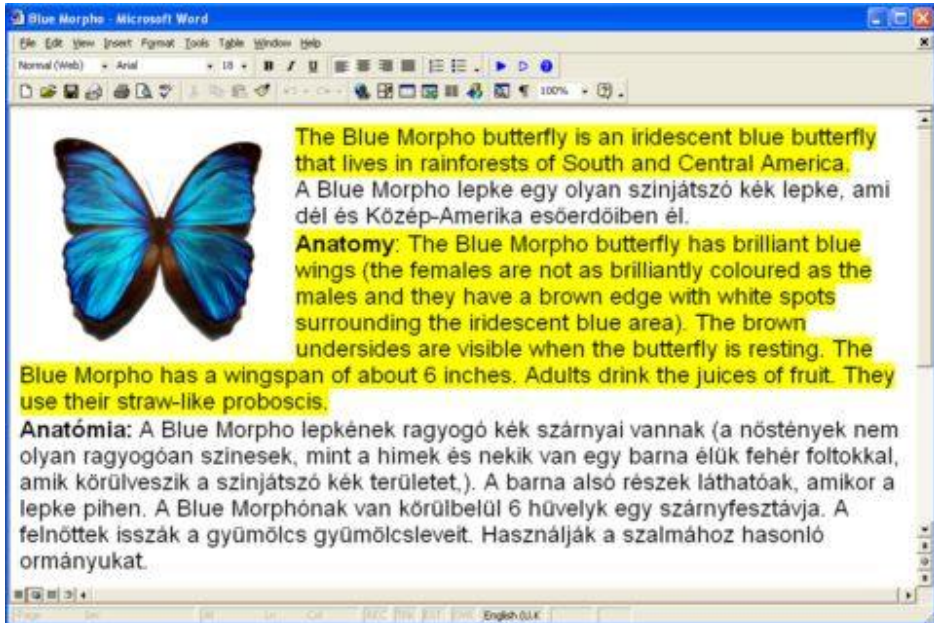
Befejeződik a hardverkulcsos védelem fejlesztése.

Lezáródik a RuBi (RuleBuilder) szabálybővítő fejlesztése. A felületet az aktuális szabály típusától/szófajától függően (itt főnév) egy másik panel egészíti ki, ahol további lexikális jegyek állíthatók be. A program TMX fájlok importjára is képes

1.9 Szeptember

Megszervezzük a CD-gyártást és az értékesítést. Az interneten publikáljuk a programmal kapcsolatos információkat. Véglegesülnek a szoftvervédelmi megoldások. Elkészül a MorphoWord termék.

A MorphoWord hagyományos CD-n kapható dobozos program a Microsoft Word szövegszerkesztőbe integrálódik. Már nem szolgáltatásként, hanem hagyományos dobozos terméként jelent meg. Két változatban készült el, az egyfelhasználós alapverzió mellett született egy intézményeknek szánt hálózatos megoldás is, amelyhez bővíthető adatbázis tartozik.



The Blue Morpho butterfly is an iridescent blue butterfly that lives in rainforests of South and Central America. A Blue Morpho lepké egy olyan színjátszó kék lepke, ami dél és Közép-Amerika esőerdőiben él.

Anatomy: The Blue Morpho butterfly has brilliant blue wings (the females are not as brilliantly coloured as the males and they have a brown edge with white spots surrounding the iridescent blue area). The brown undersides are visible when the butterfly is resting. The **Anatómia:** A Blue Morpho lepkének ragyogó kék szárnyai vannak (a nőstények nem olyan ragyogóan színesek, mint a hímek és nekik van egy barna élük fehér foltokkal, amik körülveszik a színjátszó kék területet.). A barna alsó részek láthatóak, amikor a lepke pihen. A Blue Morphónak van körülbelül 6 hüvelyk egy szárnyfesztávja. A felnőttek isszák a gyümölcs gyümölcsleveit. Használják a szalmához hasonló ormányukat.

MetaMorpho RuBi - user@localhost

Felszíni alakok

Forrás: water installation
Cél: vízellátó bekötés

Szabály típusa: főnév Szaknyelv: műszaki

Fordítás Szabály létrehozása Szabály mentése

Szaknyelvek Karbantartás Súgó

A szabály felvétele előtti fordítások:
vízellátó bekötés

A szabály felvétele utáni fordítások:

Útolsó vissza Összes vissza Kilépés

Lexikális tulajdonságok

élelőny
 nem igen

személy
 nem igen

megszámálható
 bármelyik igen nem

típus
 köznévf tulajdonnévf

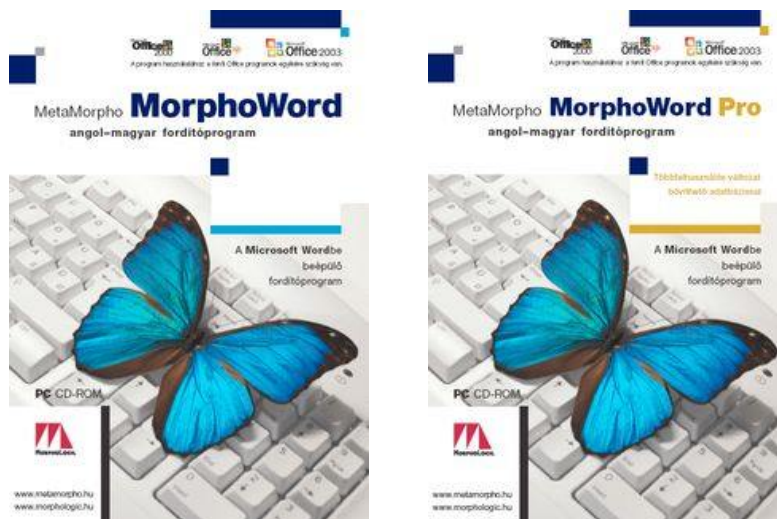
földrajzi névf
 nem igen

intézménynévf
 nem igen

nak/nek
 nem igen

10 Október

Október 21-én megkezdődik a MorphoWord fordítóprogram piaci értékesítése. A MorphoWord Pro változat nyelvi tudása az alapváltozattal azonos, de bővíthető, hálózatosan is használható és két felhasználói jogot tartalmaz.



2 Nyelvészeti fejlesztések

Az idei évben a nyelvészeti munka elsősorban a magyar-angol fejlesztésekre irányult. 2005 elején indult el a magyar-angol fordítóprogram projekt az MTA Nyelvtudományi Intézetével és a Szegedi Tudományegyetem Informatikai Tanszékcsoportjával közös együttműködésben. A fejlesztésekre az NKFP-2/008/2004 pályázat biztosítja a megfelelő kereteket. Az év eredménye, hogy meglévő eszközeinket és erőforrásainkat a projektbe importáltuk, és jelentős lépéseket tettünk abba az irányba, hogy ezekből egy valóban jól működő, robusztus fordítót hozunk létre. A munkamegosztásban a magnyelvtan és a szabályok felcímkezését végző konverter program fejlesztése, valamint a morfológia karbantartása maradt a MorphoLogicnál, az igei vonzatkeretek fordítása és a működtetéséhez szükséges felcímkezése a Nyelvtudományi Intézet, a névszói szerkezetek fordítása és szemantikai jegyekkel történő ellátása az SZTE feladata lett.

A pályázat a következő táblázatban összefoglalt két munkaszakaszt zárta le ebben az évben:

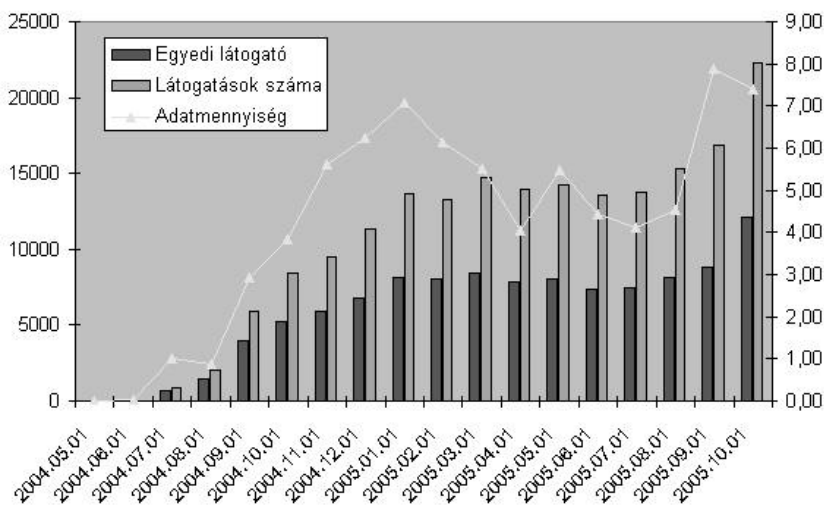
A feladat megnevezése	Közreműködő szervezetek
Tervezés (2005.01.01 – 2005.04.30)	
A rendszer architektúra specifikációja	ML, SZTE, NYT
A nyelvtan reprezentációjának specifikációja	ML, SZTE, NYT
A rendszer kiértékelési módszertanának kidolgozása	SZTE, NYT
A rendszer monotonitásával kapcsolatos kutatás	ML
Az internetes frissítések lehetőségének kidolgozása	ML
A munkaszervezés módszerének kidolgozása	ML
A teszterv kidolgozása	SZTE
Magnyelvtan megírása (2005.05. 01. – 2005.10.31.)	
A szabályok leködölása	ML
A nyelvtan pontosságának kimérése	ML, SZTE
Jegyzőkönyv és ennek alapján a szabályok kiegészítése	NYT
Újabb mérés	SZTE

A magyar-angol program nyelvészeti munkálataira ez a dolgozat nem tér ki, erről Merényi Csaba ugyanebben a kötetben található cikkében lehet olvasni. [3].

Az angol-magyar fejlesztések a MoBiCAT és a MorphoWord felhasználóinak visszajelzései alapján folytak.

3 A fordítóprogram első évének értékelése

A MoBiCAT-szolgáltatás első évét a felhasználók számának függvényében és a használati szokásokat kutató kérdőívre adott válaszok alapján értékelhetjük. A MoBiCAT ma egy szűkebb, de a visszajelzések alapján elégedett felhasználói körrel rendelkezik. A kezdeti felfutás az idei év közben stabilizálódott. Az év végén megjelenő MorphoWord a MoBiCAT-eladásokra pozitív hatást gyakorolt:



4. Tervek röviden

Jövőre el akarjuk érni a magyar szövegeken általában érthető angol fordítást produkáló nyelvészeti leírást. Javítani szeretnénk az angol magyar fordítások nyelvi minőségét az elő- és utófeldolgozás lehetőségének kialakításával (ismeretlen szavak, terminológiák) Ezt szolgálná a tervezett angol–magyar jelentésegyértelműsítő adatbázis bővítése a leggyakoribb többértelmű angol szavakkal. Szeretnénk integrálni a fordítóprogramot fordítómemóriába. Továbbfejlesztjük a nyelvi hasonlóság elvén működő intelligens fordítómemóriát, és ezt is fordítómemória-keretprogramba integráljuk.

Referenciák

1. Tihanyi László: A MetaMorpho projekt története. *I. MSzNy, Szeged (2003)*
2. Tihanyi László: A MetaMorpho projekt 2004-ben. *II. MSzNy, Szeged (2004)*
3. Merényi Csaba: A MetaMorpho magyar-angol gépi fordító rendszer igei vonatkozatait működtető nyelvten *III. MSzNy, Szeged (2005)*